MIGNEX

MIGNEX Handbook Chapter 3

Data management plan

Jørgen Carling
Peace Research Institute Oslo

Mathias Czaika
Danube University Krems

Anne Duquenne
Peace Research Institute Oslo

Marta Bivand Erdal
Peace Research Institute Oslo

Jessica Hagen-Zanker
Overseas Development Inst.

Mira Ivanova
Peace research Institute Oslo

Carlos Vargas-Silva
University of Oxford

MIGNEX deliverable D1.2 March 2020 (Version 3)

Suggested citation

Carling J., Czaika M., Duquenne A., Erdal M.B., Hagen-Zanker J., Ivanova M. and Vargas-Silva C. (2020) *Data management plan*. MIGNEX Handbook Chapter 3 (v3). Oslo: Peace Research Institute Oslo. Available at: www.mignex.org/d012.

MIGNEX

MIGNEX (Aligning Migration Management and the Migration-Development Nexus) is a five-year research project (2018–2023) with the core ambition of creating new knowledge on migration, development and policy. It is carried out by a consortium of nine partners in Europe, Africa and Asia: the Peace Research Institute Oslo (coordinator), Danube University Krems, University of Ghana, Koç University, Lahore University of Management Sciences, Maastricht University, the Overseas Development Institute, the University of Oxford and Samuel Hall.

See www.mignex.org.

MIGNEX has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 770453.

The MIGNEX Handbook

The MIGNEX Handbook grows chapter by chapter over the lifetime of the project. It is primarily as a tool for internal information-sharing and quality assurance. The text refers to 'we' as the team members and 'you' as an individual team member reader. The handbook is public in order to ensure transparency and facilitate knowledge exchange also on issues such as project management, methodology and communication.

Acknowledgements

This document was initially drafted by the lead author and critically reviewed and added to by the co-authors. The authors are also grateful to Jason Gagnon (OECD) for valuable comments. The content of the document, including opinions expressed and any remaining errors, is the responsibility of the authors.

Publication information

This work is licensed under the Creative Commons CC BY NC 4.0 License. You are free to share and adapt the material if you include proper attribution (see suggested citation), indicate if changes were made, and do not use or adapt the material in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes.

Peace Research Institute Oslo, Oslo, Norway

March 2020 (Version 3)

ISBN (print): 978-82-7288-959-2

ISBN (online): 978-82-7288-960-8

ISBN numbers refer to the first published version.

The views presented are those of the author(s) and do not necessarily represent the views of the institutions with which they are affiliated.

History of changes

Version	Date	Changes
1	28 February 2019	Version submitted as official deliverable to the EC.
2	13 March 2019	Section 3.7 (File management and naming conventions) has been updated: a sub-folder called 8-MIGNEX-resources has been added to the MIGNEX information folder structure; the sub-folders previously named 'work in progress' have been renamed 'supporting files' and the description has been changed accordingly. Several errors in the examples of file names have been corrected.
3	31 March 2020	Version re-submitted to the EC in response to the October 2019 ethics check, with correct references to GDPR legislation. Specifications have been added to various sections, reflecting the progress of the project.

Contents

Tables

Text boxes

MIGNEX Handbook Chapter

3. Data management plan

MIGNEX will collect and analyse a large amount of data in different forms. Well-planned data management is essential for using our resources efficiently, achieving our analytical objectives, and making the data a valuable part of our output.

—

We will produce four data sets, of which at least three will be openly archived.

—

Our data security measures seek to minimize the likelihood and consequences of (1) unauthorized data access and (2) data corruption or loss.

—

We make our work more efficient and secure by following a detailed and logical system for organizing and naming files and folders.

3.1 Introduction

This data management plan opens with a background to the identification of research data within the project and outlines the division of labour for further development of the plan. From section 3.2 onwards, the structure is based on the Horizon 2020 template for data management plans and the recommendations of the Consortium of European Social Science Data Archives (CESSDA).

3.1.1 What are data?

There is no consensus definition on what constitutes 'research data' and the understanding of the term differs between disciplines. In social sciences – and especially within qualitative approaches – the lines between data, analysis, and publications are not always clear-cut. The data management plan for a project such as MIGNEX should therefore include consideration and justification of what is regarded as data. We do so with three foundations.

- Research data can be defined as 'information that has been structured by methodology for the purpose of producing new knowledge'.[1] This is a definition we find useful.
- Systematic structuring and documentation are required to turn information into a data set. These requirements can go beyond the needs of the project's own workflows. The question 'what are data' is therefore also a matter of resources and priorities.
- The Description of Action (DoA) indicates what will be treated as data in the project. As with all other aspects of the project, we take the DoA as the starting point and deviate from it only if it is scientifically and financially justifiable.

From these foundations we identify *four data sets* (Table 1) We exclude other bodies of information from being treated as data sets. These other bodies of information still play important roles in project workflows and will, where appropriate, be systematically stored. However, they will not be structured and documented to the extent required to qualify as a data set. Table 1 provides justification for each exclusion.

The sections of this data management plan that concern data security, file management and naming conventions also apply to items that are not considered data sets in their own right.

Table 1. Classification of project information as research data or not

WP	Potential data	Data set	Name of data set or justification for exclusion
WP 2	Truth Table	Yes	*MIGNEX Research Area Truth Table*
WP3	Survey data	Yes	*MIGNEX Survey Data Set*
WP4	Key informant interviews	No	The information from these interviews are not recorded, structured or documented for use by others. They serve as input for the interviewing researcher to write research area interim reports. Researchers archive notes in a secure place for potential follow-up.
	Focus groups	Yes	*MIGNEX Focus Group Data Set*
	Research area coding sheets	No	These are draft products that mediate between the raw data and the research area truth table. Some of the information contained in these sheets should be part of the documentation of the research area truth table. The Research area coding sheets will be archived on the project server but not made publicly available.
	Research area interim reports	No	These documents are not systematically structured information, but rather draft text to be used as input to the Case Study Briefs and possibly other output. They will be archived on the project server.
WP5	Policy expert interviews	No	These interviews serve as input for the interviewing researcher to write the Background Paper drafts. Researchers are requested to archive any personal notes in a secure place for potential follow-up.
	Policy coding sheets	No	These sheets primarily contain the classification of policy that, after quality assurance, is imported into the MIGNEX policy database. Some of the information contained in these sheets should be part of the documentation of the research area truth table. The Policy coding sheets will be archived on the project server.
	Policy database	Yes	*MIGNEX Policy Database*
WP8, WP9	Policy documents	No	The analyses in WP8 and WP9 will partly draw directly on primary sources in the form of policy documents, rather than on data collected in WPs 3–5. These documents will not be structured by methodology in such a way that they represent research data.

[1] Adapted from the Social Sciences and Humanities Research Council of Canada.

3.1.2 Further development of the data management plan

The data management plan is a living document that will evolve through the course of the project. Detailed aspects of data management will be specified alongside the development of methodology and data collection. Responsibilities for follow-up of the initial data management plan are assigned to individuals based on their roles, as follows:

- Anne Duquenne/Mira Ivanova (as project manager)
- Jørgen Carling (as leader of WP2)
- Jessica Hagen-Zanker (as leader of WP3)
- Marta Bivand Erdal (as leader of WP4)
- Carlos Vargas-Silva (as leader of WP5)
- Mathias Czaika (as leader of tasks T2.2 and T2.6 related to QCA methodology)

3.2 Data summary

The data management plan contains information at two levels: the project level and the data set level. The data summary is structured accordingly.

Project level	MIGNEX data sets			
Data set level	MIGNEX survey data set	MIGNEX focus group data set	MIGNEX policy database	MIGNEX research area truth table

3.2.1 Project-level data summary

Purpose

MIGNEX is driven by one overarching objective: Contribute to more effective and coherent migration management through evidence-based understanding of the linkages between development and migration. Specific objectives address the causes and consequences of migration as well as the role of migration-related policy. Primary data is collected and analysed to meet the project objectives. The uses of the various data sets and their relation to the project objectives and deliverables are described in the DoA.

Data re-use

The project design does not rely on re-use of existing data other than as background information, for instance in the selection of research areas.

The final versions of the MIGNEX policy database and the MIGNEX research area truth table may incorporate existing data from sources outside the project.

- *Follow-up by Mathias Czaika: specify uses of existing data (if any) in the MIGNEX research area truth table. To be completed in conjunction with MIGNEX Handbook chapter 13 (D2.6).*

- *Follow-up by Carlos Vargas-Silva: specify uses of existing data (if any) in the MIGNEX policy database. To be completed in conjunction with MIGNEX Handbook chapter 11 (D5.2)*

Data utility

The data will primarily be useful for other researchers, analysts, and students with an interest in migration and/or development, either in the specific case study countries or more generally.

3.2.2 MIGNEX data sets

The project will produce four data sets, each summarized in a table:

- MIGNEX survey data set (Table 2)
- MIGNEX focus group data set (Table 3)
- MIGNEX policy database (Table 4)
- MIGNEX research area truth table (Table 5)

Table 2. Summary of the MIGNEX survey data set

Data type	Survey data
Unit of analysis	Individuals
Sample size	N = 12,000 (500 individuals in each of the 25 research areas)
Population	General population aged 18+ (upper age limit to be determined)
Data origin	Computer-Assisted Personal Interviews (CAPI) using tablets
Thematic coverage	Developed as task 3.1 (Dec 2018–Jun 2020) The foreseen thematic coverage described in the DoA has the following components: Household-level demographic and socio-economic information; Household's migration experiences and connections to migrants; Individual demographic and socio-economic information; Perceptions of possible futures and strategies; Migration aspirations, intentions or preparations; Perceptions of local and national development; Perceptions of migration infrastructure.
Detailed documentation	MIGNEX Handbook Chapter 7 (D3.1, due Jun 2020) provides the detailed methodology including sampling strategies, questionnaire, and quality assurance mechanisms. MIGNEX Handbook Chapter 10 (D3.2, due Jun 2021) will document the data collection process and describe the data collected, including data set variables.
Data files (after pooling)	One .dta file (Stata); corresponding files in other formats if needed Data sets in other formats will be provided if needed for analysis (e.g. household and individual level separated; country level files)
Supplementary files	.do files documenting cleaning variable generation and merging MIGNEX Handbook Chapters 7 and 10 (.pdf files)
Data set size (Mb)	Not yet known
Data set manager	Jessica Hagen-Zanker (WP3 leader)

Table 3. Summary of the MIGNEX focus group data set

Data type	Interview data
Unit of analysis	The data is organized by focus groups, but each group does not necessarily represent a unit of analysis for analytical purposes.
Sample size	2–4 focus groups in each of the 25 research areas (50–100 in total)
Population	General population aged 18–39, divided between individuals with and without personal migration experiences and connections (interviewed in separate groups).
Data origin	Audio-recorded focus group interviews
Thematic coverage	To be developed as task 4.1 (Dec 2018–Jun 2020) The foreseen thematic coverage described in the DoA has the following components: Local migration dynamics, including perceived changes over time; Perceptions of the risks and benefits of migration, with attention to gender dimensions; Perceived development impacts of migration and transnational practices; The perceived impacts of specific developments on migration dynamics; Mobility aspirations fulfilled and thwarted.
Detailed documentation	MIGNEX Handbook Chapter 8 (D4.1, due Jun 2020) will provide the detailed methodology including focus group interview guides and instructions for transcription and/or note-taking. MIGNEX Handbook Chapter 11 (D4.2, due Jun 2021) will document the data collection process and describe the data collected.
Data files (after pooling)	Not yet known (see follow-up requirements below).
Supplementary files	MIGNEX Handbook Chapters MHC8 and MHC11 (.pdf files)
Data set size (Mb)	Not yet known
Data set manager	Marta Bivand Erdal (WP4 leader)

- *Follow-up by Marta Bivand Erdal: Decide on and describe procedures for transcribing and/or taking notes from focus group interviews to ensure that relevant information is compiled in a cost-efficient way, as well as in line with the GDPR Regulation 2016/679. (This should reflect discussions in the WP2 workshop on 14 November 2018, which highlighted strengths and weaknesses of different options). Decide on and describe the subsequent approach for storage and analysis, including specification of file types and database setup, if relevant. To be completed in conjunction with MIGNEX Handbook Chapter 8 (D4.1) at the latest.*

Table 4. Summary of the MIGNEX policy database

Data type	Simple database
Unit of analysis	Countries
Sample size	N = 10 (countries)
Data origin	Primary data collection through interviews with key informants and review of policy documents; possibly also analyses or compilation of secondary data.
Thematic coverage	To be developed as task 5.1 (Dec 2018–Jan 2020) The foreseen thematic coverage described in the DoA is given by the following questions: How have major migration-relevant policies evolved over the past five years?; What types of migration-relevant policy interventions have been deployed by international bodies?; What have been the (perceived) effects of these policy interventions on migration?; How has migration affected the national or regional development policy context?; How do policies directly or indirectly affect transit migration?
Detailed documentation	MIGNEX Handbook Chapter 9 (D4.1, due Jan 2020) will provide the detailed methodology including key informant interview guides. MIGNEX Handbook Chapter 12 (D5.2, due Jun 2021) will document the data collection process and describe the data collected.
Data files (after pooling)	Not yet known (see follow-up requirements below).
Supplementary files	Not yet known. For documentation of the data, it may be necessary to extract relevant parts of MHC9 and MHC12 since those chapters also describe other aspects of the policy review that are not directly related to the policy database.
Data set size (Mb)	Not yet known
Data set manager	Carlos Vargas-Silva (WP5 leader)

- *Follow-up by Carlos Vargas-Silva: Decide on and describe the format of the MIGNEX policy database, including the specification of file type(s). Consider how the background for classifications will be documented and stored, i.e. whether the policy coding sheets will be publicly available in original or modified form, and whether the database will contain any information beside the final classifications. To be completed in conjunction with MIGNEX Handbook Chapter 9 (D5.1).*

Table 5. Summary of the MIGNEX research area truth table

Data type	Table set up for qualitative comparative analysis (QCA)
Unit of analysis	Research areas
Sample size	N = 25 (research areas)
Data origin	Systematic compilation and classification of information from the MIGNEX survey data set, the MIGNEX policy database, and qualitative fieldwork in each research area.
Thematic coverage	To be developed as task 2.6 (Jun 2019–Dec 2019) The thematic coverage will reflect the information available from the data origins listed above. The data set will contain all the conditions that are documented and may be included in the analyses, not exclusively the conditions that are eventually used in the analyses in MIGNEX publications.
Detailed documentation	MIGNEX Handbook Chapter 6 (D2.5, due Dec 2019) will provide the detailed specification of conditions to be included. MIGNEX Handbook Chapter 13 (D2.6, due Jun 2021) will document the process of compiling the research area truth table and describe the content of the final table.
Data files (after pooling)	Not yet known (see follow-up requirements below).
Supplementary files	MIGNEX Handbook Chapters 6 and 13 (.pdf files).
Data set size (Mb)	Not yet known
Data set manager	Mathias Czaika (Task leader of T2.2 and T2.5)

Note: 'Truth table' is the standard name of data tables for qualitative comparative analysis. Each variable in the table is known as a 'condition'.

- *Follow-up by Mathias Czaika: Decide on and describe the format of the MIGNEX research area truth table, including the specification of file type(s). Consider how the background for classifications will be documented and stored, including whether the database will contain any information beside the final classifications. To be completed in conjunction with MIGNEX Handbook Chapter 6 (D2.5).*

3.3 FAIR data

MIGNEX adheres to the principle of making data 'FAIR', which means ensuring that it is *Findable*, *Accessible*, *Interoperable*, and *Re-usable* (Wilkinson et al. 2016). The following sections describe how we will ensure compliance with each of these ideals.

3.3.1 Making data findable

The data sets will be deposited with the Norwegian Centre for Research Data (NSD). The Coordinator has established contact with NSD and made enquiries regarding data archiving and the development of new functionalities in the coming years. The data management plan thus reflects the expected capacities of NSD at the time of archiving. NSD is a member of the Consortium of European Social Science Data Archives (CESSDA) and the data sets will be included in the CESSDA Data Catalogue.

Each of the data sets will have a Digital Object Identifier (DOI). It will be possible to deposit the data and have a DOI assigned before the data are released for public use. In this way, MIGNEX publications that make use of the data while they are still under embargo can include a stable identifier that links to the openly archived data sets.

The approach to *naming*, *keywords*, *versioning* and *metadata creation* will comply with the standards of NSD. Good practice for file naming and version management is implemented from an early stage in the project, before data collection begins. Procedures for file management and naming are described in detail in section 3.7.

3.3.2 Making data openly accessible

MIGNEX survey data set, MIGNEX policy database and MIGNEX research area truth table

The MIGNEX survey data set, MIGNEX policy database and MIGNEX research area truth table will be made openly available through NSD and CESSDA. The associated metadata, documentation and code (when relevant) will be deposited together with the data sets. The data will be deposited in standard file types conforming to the NSD file type recommendations.

- *Follow-up by Jessica Hagen-Zanker: Specify the file type(s) used in the MIGNEX survey data set. To be completed in conjunction with MIGNEX Handbook chapter 7 (D3.1)*
- *Follow-up by Carlos Vargas-Silva: Specify the file type(s) used in the MIGNEX policy database. To be completed in conjunction with MIGNEX Handbook chapter 9 (D5.1)*
- *Follow-up by Mathias Czaika: Specify the file type(s) used in the MIGNEX research area truth table. To be completed in conjunction with MIGNEX Handbook chapter 6 (D2.5)*

MIGNEX focus group data

The MIGNEX focus group data set potentially contains person-identifiable data. Access to this data set is determined based on the principle *as open as possible, as closed as necessary*. Recordings are selectively transcribed to remove indirectly identifiable information. This is only realistic because conversations focus primarily on community-level issues and not on personal experience. The de-identified data set is archived openly for re-use. The associated metadata, and documentation will be deposited together with the data set.

- *Follow-up by Marta Bivand Erdal: Verify that the degree of identifiability in the data is consistent with the subsequent storage of the data set. To be completed in conjunction with MIGNEX Handbook chapter 12 (D4.2).*

3.3.3 Making data interoperable

Basic interoperability for all four data sets is ensured using standard file types, adherence to naming conventions, and provision of detailed documentation, in line with the guidelines of NSD and CESSDA. As an aspect of interoperability, we also consider possibilities for extending each data set with additional cases or variables, and for making comparisons with other data.

MIGNEX survey data set

The MIGNEX survey data set will be archived as a .dta file for Stata. Although Stata is proprietary software, the file type can be converted to other formats by end users, depending of their preference.

The questionnaire will draw on existing questions and classifications where possible and relevant. These could include questions on subjective well-being and classifications on occupations, for instance.

- *Follow-up by Jessica Hagen-Zanker: Consider interoperability in the design of the questionnaire. To be completed in conjunction with MIGNEX Handbook chapter 7 (D3.1).*

MIGNEX focus group data

The data file type for the MIGNEX focus group data set is yet to be determined. If the data set takes the form of an NVivo database, the raw data will also be archived in standard file types conforming to the NSD file type recommendations (e.g. Rich text format). See 3.3.2 on possible access restrictions.

- *Follow-up by Marta Bivand Erdal: Specify the file type of the MIGNEX focus group data set once it has been decided. To be completed in conjunction with MIGNEX Handbook Chapter 8 (D4.1).*

Since the MIGNEX focus group data set will not rely on structured questions and responses, interoperability-enhancing use of standard classifications or vocabularies is not relevant.

MIGNEX policy database

The MIGNEX policy database will be archived in a standard file, yet to be determined.

The database will have a high level of basic interoperability by using countries as the unit of analysis. It can easily be extended with other country-level data for the same countries. Throughout the project, country-level data will be identified by ISO 3166-1 alpha-3 three-letter country codes (see 3.7.4). Classifications from existing databases will be adopted where possible and relevant.

- *Follow-up by Carlos Vargas-Silva: Specify the file type(s) to be used in the MIGNEX policy database. To be completed in conjunction with MIGNEX Handbook Chapter 9 (D5.1).*
- *Follow-up by Carlos Vargas-Silva: Consider interoperability in the specification of the MIGNEX policy database. To be completed in conjunction with MIGNEX Handbook Chapter 9 (D5.1).*

MIGNEX research area truth table

The MIGNEX research area truth table will be archived as a standard file, yet to be determined.

- *Follow-up by Mathias Czaika: Specify the file type(s) to be used in the MIGNEX research area truth table. To be completed in conjunction with MIGNEX Handbook chapter 6 (D2.5).*
- *Follow-up by Mathias Czaika: Consider interoperability in the specification of the MIGNEX research area truth table, e.g. whether there are standards or approaches to coding fuzziness that should be adhered to. To be completed in conjunction with MIGNEX Handbook chapter 6 (D2.5).*

3.3.4 Increasing data re-use

Licensing

All the MIGNEX data that are archived with NSD will have a Creative Commons CC BY 4.0 license. This asserts that users external to the project may copy, distribute, use and adapt the data (also for commercial purposes) but only if they include proper attribution. (See 3.5.1 on the attribution of authorship for data sets).

Embargo

MIGNEX data sets will be released for public use as early as possible, with the limitations that (1) data integrity, data protection and data documentation must be first assured, and (2) the

project teams' own analyses of the data must first be published. (This only refers to analyses identified as tasks in the project and taking place during the project period.)

Based on these principles, the planned release date for the MIGNEX data sets is during the final month of the project (August 2023). The embargo until this time will ensure that the first public release of project results is in the form of project publications.

The Steering Committee may decide on an *earlier release* of a data set with the approval of the data set authors and the leaders of tasks involving analyses of the data set.

The Steering Committee may decide on a *later release* of a data set if there are unresolved issues relating to data quality or data protection.

Re-usability

The MIGNEX data sets will be re-usable by third parties, with possible exceptions for the MIGNEX focus group data set (see 3.3.2). Measures to facilitate extensions and comparisons are described in 3.3.3).

Durability

The data will remain re-usable for an indefinite period. Their analytical value will change as time passes and empirical realities change. The data will eventually become most valuable as historical documentation.

Quality assurance processes

Data quality is central to increasing data re-use, and it is therefore considered an element of FAIR data. Quality assurance processes are therefore included in this section of the standard template for data management plans. However, the specific precautionary measures are integral to the methodology and protocols of work packages 6–9 and will therefore be developed in the relevant MIGNEX Handbook Chapters:

- 6. QCA conditions and measurement (D2.5)
- 7. Survey data collection (D3.1)
- 8. Qualitative data collection (D4.1)
- 9. Country-level policy review (D5.1)

Each of these chapters will identify data quality risks and describe precautionary measures. Data quality is *the degree to which data meet the purposes and requirements of their use.*[2] The types of risks to be addressed therefore occur when data

- do not fit the project's analytical objectives as well as possible
- are collected in ways that fails to capture important empirical variation
- are marked by human error, for instance in translation, data entry, or mishandling of files
- suffer from poor documentation.

Promotion

Re-use of the data sets will be increased through active promotion of the data sets around the time of their release. Possible actions include the following:

[2] Definition circulated by UK data service and others in the context of the 2017 of 'Love Your Data week', an international celebration of all things data related.

- Submission of data articles to relevant open-access journals (e.g. *Demographic Research*)
- Submission of posts to relevant blogs (e.g. the OECD *Development Matters* blog)
- Publication of supplementary resources for using the MIGNEX data sets in teaching
- Dissemination of information about the data sets to the End User Panel

3.4 Allocation of resources

The costs of making data FAIR are estimated to be relatively low. Relevant standards, naming conventions, quality assurance and documentation are integrated parts of managing data for the project's own use. Data archiving with the NSD is free of charge. Long-term preservation will not result in additional costs. *Curation and documentation of project data* is identified as task T10.7 (as an aspect of impact maximization) and allocated 1.3 person-months in the budget.

Data management responsibilities within the project are shown in Table 6.

Table 6. Distribution of data management responsibilities

Data set	Ensuring data security (see section 3.5)[1]	Quality-assuring data sets and documentation (see section 3.3.4)	Making data FAIR (see section 3.3)
MIGNEX data overall	Anne Duquenne/Mira Ivanova (for posted data sets)	Jørgen Carling (coordination)	Jørgen Carling (coordination)
MIGNEX survey data set	Jessica Hagen-Zanker (during data preparation)	Jessica Hagen-Zanker	Jessica Hagen-Zanker
MIGNEX focus group data	Marta Bivand Erdal (during data preparation)	Marta Bivand Erdal	Marta Bivand Erdal
MIGNEX policy database	Carlos Vargas-Silva (during data preparation)	Carlos Vargas-Silva	Carlos Vargas-Silva
MIGNEX research area truth table	Mathias Czaika (during data preparation)	Mathias Czaika	Mathias Czaika

Note: (1) 'Data preparation' includes collection, initial storage, pooling and processing, as relevant to each data set.

3.5 Ethical aspects

3.5.1 Data authorship

The archiving of data sets with DOIs and metadata require the identification of data set authors. Our existing authorship guidelines (included in the Consortium Agreement and MIGNEX Handbook Chapter 5) only address authorship of publications.

It is increasingly common to archive data for public use and formally identify data authors, but there are no established criteria or guidelines for authorship attribution. A recent article on the issue included the following proposal:

> We propose that, in order to be cited as a data author, a person must have made substantial contributions to the original acquisition, quality control, and curation of the data, be accountable for all aspects of the accuracy and integrity of the data provided, and ensure that the available data set follows FAIR Guiding Principles, which instruct that the data and metadata meet criteria of findability, accessibility, interoperability, and reusability (Bierer *et al.* 2017:1685).

Based on this proposal and our publication authorship guidelines, MIGNEX adopts the following guidelines for data set authorship.

Data authorship guidelines

1. Data authorship is assigned at the level of each data set in the form that it takes for archiving after the end of the project (e.g. pooled from various locations).
2. Authorship is formally established when the data set is assigned a Digital Object Identifier (DOI) by a data archive.
3. Everyone who is listed as an author should have made a substantial direct academic contribution (i.e. intellectual responsibility and substantive work) to at least two of the following four components:

   - Development of specifications for data collection (e.g. questionnaire design)
   - Collection of raw data (e.g. interviewing)
   - Initial processing and/or preparation of the data (e.g. transcription, data cleaning)
   - Curation of the data, including provision of metadata and documentation

4. The data authors share responsibility and credit for the quality and integrity of the data but are not responsible for the conclusions drawn from subsequent analyses unless they are also listed as authors of the relevant publications.
5. If one person has had overall responsibility for the creation of the data set, that person is entitled to be the first author. All other authors should subsequently be listed alphabetically.

3.5.2 Ethical issues in data collection and processing

Ethical issues in data collection and processing are addressed in MIGNEX Handbook Chapter 4, *Research ethics and research integrity* (D1.3).

To the extent to which it applies, the data sets are also subject to the GDPR Regulation 2016/679, with methodological design including measures to ensure that the rights and freedoms of the research participants are safeguarded, and all data protection measures are GDPR compliant.

3.6 Data security

Data security is concerned with mitigating two risks:

- Unauthorized data access
- Data corruption or loss

Each risk takes specific and different forms through the life cycle of data and needs to be managed accordingly. Data security must be ensured not only for the information that is curated and eventually archived as data sets, but *for all the information listed as 'potential data'* in Table 1.

3.6.1 Unauthorized data access

The meaning of unauthorized data access requires clarity about the conditions for *authorized* access. In MIGNEX, access to data is authorized if it is covered by one of these conditions:

Necessary access for acquisition and processing Interviewers, enumerators, field supervisors, transcribers and other personnel who are not MIGNEX team members will engage directly with the data as it is acquired and processed. This is a specific and time-limited need

and access will be revoked when it is no longer necessary. In the case of subcontracting, data processing by non-team members is governed by data-processing agreements between the consortium members as data controllers and the respective subcontractor, the data processor.

Necessary or valuable access for analytical purposes MIGNEX team members will need access to the data for analysis. This is not only a direct and specific need; it benefits the analysis that all team members have access to as much of the data as possible during the lifetime of the project. All team members will therefore have access to data that is posted in the MIGNEX data folder (see 3.7.1). Access is revoked for members who leave the MIGNEX team.

Unavoidable access in the context of data collection Survey interviews and key informant interviews may take place in the presence of third persons who will have access to the raw data through hearing what is said. It is highly context-dependent whether it is feasible or socially acceptable to prevent the presence of third persons. Such unavoidable access is regarded as authorized when the interviewee and the interviewer are aware of it and deem it acceptable. If an interview is overheard without the awareness of the interviewee and the interviewer, it is a case of unauthorized data access. (This is also an ethical and methodological issue and is also addressed in MIGNEX Handbook Chapters 4 and 7.)

The main risks of unauthorized data access, when data are collected through fieldwork, can be differentiated by type of storage.

Unprotected storage devices (notebooks, recorders, unencrypted pen drives) can be the source of unauthorized data access if they are lost, stolen, left unattended, or their content is not destroyed or deleted after use.

Protected storage devices (digital storage with access restrictions set by user profiles or encryption keys) can be the source of unauthorized data access when, for instance, credentials or keys are revealed, when access restrictions are breached through targeted attempts, or when access rights are not granted and revoked according to the protocol.

3.6.2 Data corruption or loss

Data can be corrupted or lost when the storage device is lost, stolen, destroyed or malfunctions. Digitally stored data can also be lost or corrupted through human error such as accidental deletion, overwriting or editing.

3.6.3 Data security measures

Much of the MIGNEX data will be acquired and initially stored in a dispersed way before it is stored and managed centrally on the project's cloud server, MIGNEX OneDrive. The distinction between dispersed and central storage structures our data security measures.

In general terms, risk is the combined product of the *likelihood* and the *consequences* of an adverse event. Our approach to mitigating data security risks therefore consist of measures that minimize (1) the likelihood and (2) the consequences, respectively, of unauthorized data access and data corruption or loss. Table 7 Presents this structure, lists specific objectives, and identifies which ones are relevant to central and dispersed storage. The following sections of the text address each of the objectives for mitigating data security risks in each type of storage.

Table 7. Strategic objectives for mitigating data security risks

	Unauthorized data access	Data corruption or loss
Minimizing the likelihood	• Minimizing the use of unprotected storage devices (D) • Protecting storage devices against theft or loss (D) • Granting and revoking access according to need (C) • Deleting/destroying data copies after use (D)	• Protecting storage devices against theft or loss (D) • Granting reading rights rather than editing rights when possible (C) • Following systematic procedures for managing and naming files (C, D)
Minimizing the consequences	• Minimizing the storage of personally identifiable data (C, D)	• Ensuring that data is backed up frequently and reliably (C, D)

Note: C and D indicate relevance for central and dispersed storage, respectively.

3.6.4 Central storage (MIGNEX OneDrive)

As early as possible in the process of acquiring data, the data will be stored on MIGNEX OneDrive. The folder structure and storage procedures for MIGNEX OneDrive are described in section 3.7.4.

Granting and revoking access according to need.

Access to files on MIGNEX OneDrive will be set at the level of folders and individuals. The principles for who should have access to what are detailed in section 3.7.5.

Only people who are assigned to the item are able to access it. Both folders and files can be assigned two levels of permission: 'Can view' and 'Can edit'. In both cases, it is possible to copy or download the files, but only those with 'can edit' status will be able to make the changes to the file or folder. It is possible to stop sharing an item or change the permissions at any time, or to attribute a link expiration date.

The access rights to each file and folder is managed by the Project Manager, based on the overview of credentials in the MIGNEX-team excel file.

Granting reading rights rather than editing rights when possible

Each new version of data and publication files are 'posted' as explained in section 3.7.1. The entire team has reading rights to these files but only a small number of managers have editing rights. This procedure reduces the likelihood of accidental data corruption or loss.

Following systematic procedures for managing and naming files

We have systematic procedures for naming, saving and posting files, detailed in section 3.7. The procedures include clear versioning of files.

Minimizing the storage of personally identifiable data.

See MIGNEX Handbook Chapter 4, *Research ethics and research integrity* (D1.3).

Ensuring that data is backed up frequently and reliably

Project data and documentation are stored centrally on the project's cloud service, MIGNEX OneDrive. The MIGNEX backup routines are thus a combination of the OneDrive default settings and adopted practices ensuring the safety and sound management of data.

OneDrive is hosted in Sharepoint Online, where information is secured as part of the service and according to data protection schemes in Microsoft Azure. In the case of accidental deletion, Sharepoint Online has a manually adjustable file retention period for data after its deletion. MIGNEX OneDrive uses the default setting for this period, which is 93 days from the time the file is deleted from its original location.

Earlier versions of a file remain restorable as long as the file exists on OneDrive. To edit sensitive data out of a file, the file itself needs to be permanently removed from the cloud storage to ensure no earlier versions of the file are retained.[3]

3.6.5 Dispersed storage

Dispersed storage of data will occur in work packages 3–5 which involve data collection in countries of origin and transit.

MIGNEX Handbook Chapters 7 (D3.1), 8 (D4.1) and 9 (D5.1) will include workflow for data management from the moment it is collected until it is posted as final data sets. The workflows will cover the following plus additional ones as required in each case:

- How data are stored as they are being collected
- How data are initially stored and backed up in the field, e.g. at the end of each day
- How and when raw data are transferred to the MIGNEX collaboration folder
- How data files are organized while the raw data are being processed

The workflows will include provisions for data security at each stage, following the risk mitigation objectives described in general terms in this section.

▶ *Follow-up by Jessica Hagen-Zanker, Marta Bivand Erdal and Carlos Vargas-Silva: develop the workflows described above. To be completed in conjunction with MIGNEX Handbook Chapters 7 (D3.1), 8 (D4.1) and 9 (D5.1). Some elements will be unique to each work package while others can be replicated across work packages.*

Minimizing the use of unprotected storage devices

Whenever possible, storage in protected locations or with protection measures such as encryption is preferred. For instance, pen drives for backup in the field should be encrypted.

Still, recording of data on unprotected devices (e.g. paper pads for interview notes) may be required in some circumstances. In such cases, two precautions are taken:

- The unprotected data is kept to a minimum (e.g. by leaving out identifying information)
- The data is transferred to protected digital storage (e.g. by scanning) as soon as possible

Protecting storage devices against theft or loss

Storage devices should be locked away or be under supervision. Special precautions against theft or loss should be taken when a storage device is unprotected and/or contains the only copy of the data (e.g. handwritten notes from an interview).

Deleting or destroying data copies after use

Data on paper should be destroyed as soon as it is transferred to electronic storage. Data on portable devices should be deleted when it is copied to MIGNEX OneDrive. (Data for processing

[3] See https://docs.microsoft.com/en-us/office365/enterprise/office-365-sharepoint-online-data-deletion and https://docs.microsoft.com/en-us/office365/enterprise/office-365-dealing-with-data-corruption.

or analysis may be synced to local copies of OneDrive folders). Workflows for data management should include deletion of data and MIGNEX team members should confirm that subcontractors or assistants deleted data upon completion of their work. The deletion of data processed by subcontractors is governed by data processing agreements between the consor–tium members as data controllers and the respective subcontractor as the data processor.

Minimizing the storage of personally identifiable data.

The management of personal data is addressed in MIGNEX Handbook Chapter 4, *Research ethics and research integrity* (D1.3).

Following systematic procedures for managing and naming files

Initial storage and processing of data – in the field and/or in the MIGNEX collaboration folder – must follow procedures developed by the WP leader (or a WP contributor) and described in MIGNEX Handbook Chapters 7–9. The approach to organizing and naming files should reflect the project's general file management and naming conventions (see 3.7).

Ensuring that data is backed up frequently and reliably

Workflows for data collection that involve dispersed storage will include backup procedures, for instance to encrypted pen drives where there is no reliable internet access. The backup should be stored as separately from the original as possible.

3.7 File management and naming conventions

The following principles and procedures will help us keep our files secure and orderly – not only data files, but also work in progress and different versions of deliverables, for instance. This refers to all files, not only those that are intended for public use.

3.7.1 Posted files

Our workflows are organized around milestones in the form of *posting* files. the 'posting principle' is simple but absolute: *once a file is posted, it should never be changed.*[4] If it later needs correction or updating, it will be superseded by a newer *version* of the file.

Posted files are placed in specific folders and differentiated by their file name (see below). The remaining files fall into three categories:

- Work in progress between posted versions of a file
- Continuously updated files (e.g. overview of team members) that are never posted
- Other files (e.g. a report from outside the project)

3.7.2 Versions

Systematic versioning is a key aspect of sound data management. We follow these procedures:

- Versions are numbered with integers: 1, 2, 3… (no decimals)
- In file names, the version number is preceded by a lowercase 'v': v1, v2, v3…
- New files that will at some stage be posted are created with 'v1' in the file name
- Version 2 is created when we start work on a replacement of the posted version 1

[4] See Long (2009:125) for a discussion.

- Posted files have a 'p' for posted appended to the version number: v1p, v2p, v3p...
- Words such as 'final' or 'latest' are not used in file names to distinguish between versions

3.7.3 Dates

Our standard formats for file names include a date, assigned in the following way:

- Initially, file names include the date when the file was created
- Subsequently, files are given a new date when a new draft or version is created (see below)
- Dates in file names are written as YYYY-MM-DD (e.g. 2019-01-22)

The date format ensures that when files are sorted alphabetically, they appear in the right order. In other contexts (e.g. in text) we follow standard British formatting: 21 January 2019.

Text box 1. Examples to illustrate versioning, dating and posting

[File name]-v1-2019-02-10	*The file is created and work on version 1 starts*
[File name]-v1-2019-02-15	*Work continues, and the file is saved with a new date*
[File name]-v1p-2019-02-20	*The version is finalized and posted on 20 February*
[File name]-v2-2019-03-15	*Needs for updates emerge, and work starts on version 2*
[File name]-v2p-2019-03-20	*The version is finalized and posted on 20 March*

See section 3.7.4 and regarding the first part of the file name

3.7.4 Naming files and folders

Systematic and consistent naming of data files is essential for ensuring data integrity, managing versions correctly, and preventing data loss. This concerns not only data files but also drafts, deliverables and other documents that are part of MIGNEX workflows.

Scope

The number of people involved in MIGNEX is large, and many team members play relatively small roles. Insisting on specific naming for all project-related files may therefore be unproductive. The naming conventions for project files have the following scope:

- They *must be followed* for all files that are *posted* (see 3.7.1)
- They are *strongly recommended* for drafts that are part of MIGNEX workflows
- They are *possible to use* for all project-related files

Characters to avoid

Some of our files will require (or benefit from) computer-friendly names that can be used in links and commands. Files and folders are therefore named on the bases of these rules:

- Do not use spaces or any of the following characters: . \ / : * ? " < > | [ ] & $,

- Always use *hyphens* instead of spaces to separate words.[5]

Work package, task and deliverable numbers

In names of files and folders, work package, task and deliverable numbers include a leading zero to ensure that deliverables from WP10 are listed after those from WP9 and not WP1. In task and deliverable numbers, the '.' is omitted

Examples:	WP2 →	WP02
	WP10 →	WP10
	T1.3 →	T013
	D2.2 →	D022
	D10.1 →	D101

When one deliverable consists of several outputs, each one is given a letter after the deliverable number (e.g. D095a, D095b, D095c).

In work packages 3–5, which cover data collection in countries of origin and transit, files and folders relating to each country can be differentiated by three-letter country codes rather than task numbers (see list under 'Country codes' below).

Publication types

In internal file names, publication types are identified with standard abbreviations

MBP	MIGNEX Background Paper
MCSB	MIGNEX Case Study Brief
MHC	MIGNEX Handbook Chapter (numbered MHC01, MHC02, etc.)
MPB	MIGNEX Policy Brief
MR	MIGNEX Report

MIGNEX Handbook Chapters is the only publication type that is numbered. In file names, the number contains a leading zero to ensure that the chapters are sorted correctly.

Document titles

Long titles can be shortened in file names by including only the beginning or dropping non-essential words. Capitalize the first letter of the title.

Author names

Author names are used in the file names of publication files on the MIGNEX web site (see MHC5) and for saving external publictions. With two authors, names are separated by the word 'and'. With three or more authors, the first name is followed by 'et al'.

Examples: ...Andersson-and-Siegel-2019...
...Siegel-et-al-2019...

Reviewer initials

When files are circulated for review, reviewers add their initials after a hyphen at the end of the file name. When the same file is passed on to several reviewers, several initials are added.

[5] Technically it is also fine to separate by means of underscores (case_study_brief) or capital letters (CaseStudyBrief) but for the sake of consistency and simplicity, we always use hyphens (case-study-brief).

Example:	…v1-2019-03-18.docx	*(File sent out for review)*
	…v1-2019-03-18-NM.docx	*(File with comments by NM)*
	…v1-2019-03-18-NM-AK.docx	*(File with comments by NM and AK)*

Country codes

Countries are identified by the standard three-letter country code (ISO 3166-1 alpha-3):

AFG	Afghanistan
CPV	Cabo Verde
ETH	Ethiopia
GHA	Ghana
GIN	Guinea
NGA	Nigeria
PAK	Pakistan
SOM	Somalia
TUN	Tunisia
TUR	Turkey

Research area codes

Research areas are identified by the country code followed by 1, 2 or 3 (*e.g.* AFG1, ETH3). A list of research area names and codes will be posted in the MIGNEX information folder.

Uniqueness

Each file and folder must have a *unique name* that does not require any context. This is a precaution against misplacing files, for instance when the list of recently used folders contains a sub-folder with the same name as many other subfolders. As an example, sub-folders for deliverables should include the name of the WP.

Not unique:	…/WP02/Deliverables/
	…/WP03/Deliverables/
Unique:	…/WP02/WP02-deliverables/
	…/WP03/WP03-deliverables/

Similarly, files that exist in parallel versions for each research area, country, consortium meeting, etc. should have unique names (e.g. NGA1-interim-report…, CM2-agenda…).

The exception to the uniqueness requirement is 'Old-drafts' subfolders which contains less vital files and can all have the same name in all locations (see 3.7.5)

Naming of internal files

The most important types of internal files have a fixed naming formula.

- Deliverables in progress

  Formula: [Deliverable number]-[Publication type abbreviation]-[Title]-[Version]-[Date]

  Examples: D022-MBP-QCA-for-migration-and-development-v1-2019-03-18.docx
  D051-MHC05-Impact-maximization-v1-2019-05-10.docx

- Deliverables that are posted ('p' added after version number)

  Formula: [Deliverable number]-[Publication type abbreviation]-[Title]-[Version]p-[Date]

  Examples: D022-MBP-QCA-for-migration-and-development-v1p-2019-08-30.docx
  D051-MHC05-Impact-maximization-v1p-2019-05-30.pdf

- Initial output from data collection

Formula: [Document type]-[Country or research area code]-[Version]-[Date]

Examples: Survey-execution-report-GHA-v1-2020-03-30.pdf
Survey-data-GHA-v1-2020-03-30.dta

- Complete MIGNEX data sets

Formula: [Data set name]-[Version]['p' if posted]-[Date]

Examples: MIGNEX-survey-v1-2021-06-15.dta (before posting of the first version)
MIGNEX-survey-v1p-2021-06-20.dta (first posted version)

When the workflows in a work package include other types of files, the WP leaders should specify formulas for naming them consistently.

Naming of publication files on the MIGNEX website

The publication files for the website are named in a way that is more intuitive for end users and encourages correct citation (see MHC5).

3.7.5 Folder structure and access rights

Project-related files are stored and shared through MIGNEX OneDrive, a cloud-based server. The drive is divided in four top-level folders as shown in Table 8.

Structure of subfolders

The subfolders of the MIGNEX information folder, MIGNEX data folder and MIGNEX collaboration folder will evolve over time and be flexible, but still have a consistent logic.

- The location of subfolders reflects the differences in access rights and posting versus sharing between the three top-level folders (see Table 8)
- The organization of subfolders generally follows the organization of the project, for instance by having a subfolder for each task of a WP (as described in the DoA)
- The naming of folders and files should ensure that they appear in a logical order when they are sorted alphabetically
- In the MIGNEX collaboration folder, every WP folder has a subfolder for each task and deliverable. Other subfolders can be added according to need by any team members with editing rights to the WP folder.

Table 8. Top-level folders of MIGNEX OneDrive

Location	Type of content	Posted files	Other files	Viewing rights	Editing rights
MIGNEX information folder	Reference documents and resources (e.g. templates, MIGNEX publications, agendas and minutes)	Yes	Yes (Selected files that are continuously updated)	All team members	Designated individuals in WP1 and WP10
MIGNEX data folder	Posted files that are part of project workflows (e.g. cleaned data sets, coding files, survey reports)	Yes	None (Data files in progress are in the MIGNEX collaboration folder).	All team members	Designated individuals for each subfolder
MIGNEX collaboration folder	Files for processing, analysis and drafting (e.g. data sets undergoing cleaning, draft deliverables)	None	Yes	Designated individuals for each subfolder	Identical to viewing rights

Table 9 shows the structure of subfolders. The following sections of the text elaborate on specific issues and types of subfolders. Text box 2 provide examples.

Table 9. Folder structure in MIGNEX OneDrive

Folders and subfolders	Descriptions and comments
MIGNEX-information-folder[6]	
1-Key-reference-documents	*e.g. DoA and consortium agreement*
2-Key-management-files	*e.g. overview of tasks and deliverables*
Steering-Committee-meetings	
3-Info-e-mails	
4-MIGNEX-Handbook	*PDFs of the latest version of each chapter*
MHC-old-versions	*PDFs of previous versions of each chapter*
MHC-Word-files	*Word files for all posted versions of each chapter*
5-MIGNEX-publications	*PDFs of the latest version of each publication*
Pub-old-versions	*PDFs of previous versions of all publications*
Pub-Word-files	*Word files for all posted versions of publications*
6-Templates-and-branding	
7-MIGNEX-events	*Centrally organized events*[7]
CM1-2018-11	*Consortium meetings, with year and month*
CM2-2019-11	
…	
8-MIGNEX-resources	
MIGNEX-images	
MIGNEX-data-folder	*Posted versions of initial output*[8] *and data sets*
WP02-data	
…	*Data folders for additional work packages*

[6] Unlike the other folders, the MIGNEX information folder does not have subfolders for each WP. The subfolders are therefore numbered in order to be listed in a meaningful order.
[7] Since only designated individuals in WP1 and WP10 have editing rights to the MIGNEX information folder, files for events at the work package level are placed within the work package folders in the MIGNEX collaboration folder.
[8] All intermediate information that should be *posted*, such as survey execution reports and policy coding sheets (see Figure 3 in the DoA).

MIGNEX-collaboration-folder	
All-collab	
Other-collab	*Sub-folders for cross-cutting collaboration in groups*
Publishing	*Submission of raw files for publishing processes*
WP01-collab	
WP02-collab	*WP02 expanded to show example of content within a WP folder*
D021-drafts	*Current draft of D2.1 (empty when there is no drafting)*
Old-drafts	*Old drafts of D2.1 (see explanation below)*
D022-drafts	
Old-drafts	
D022-supporting-files	*Optional (see explanation below)*
…	
T021-Coordination	*Files related to task T2.1*
T022-Development-of-QCA	
…	
WP02-planning	*Optional: files related to WP planning, if relevant*
WP02-events	*Optional: files related to WP events, if relevant*
WP03-collab	
WP04-collab	
WP05-collab	
…	
WP10-collab	

Text box 2. Examples to illustrate the life cycle of a MIGNEX file

D013-MHC03-Ethics-v1-2019-01-10.docx	*A first version of the file is created from the template. It is identified as version 1 and labelled with the date it was created.*
D013-MHC03-Ethics-v1-2019-01-18.docx	*At regular intervals, the file is stored in 'Old-drafts' and work continues on a new draft which is given today's date.*
D013-MHC03-Ethics-v1-2019-01-18- AK.docx	*When the file is sent for review, reviewers add comments in the file and append their initials to the file name.*
D013-MHC03-Ethics-v1p-2019-02-20.docx	*When the text is approved and signed off, the document is saved as Version 1 with the current date and a 'p' appended to the version number. It is posted to the MIGNEX information folder for safe storage.*
D013-MHC03-Ethics-v1p-2019-02-20.pdf	*The PDF version is posted in the same location, with an identical file name, and uploaded to the EC website as our official deliverable.*

Old drafts

In the MIGNEX collaboration folder, folders for drafts of deliverables (e.g. D022-drafts) include a subfolder called 'Old-drafts'. This allows for keeping older drafts without having a cluttered folder. The main folder (e.g. D022-drafts) thus only contains the draft that is currently being worked on. If the deliverable is submitted and there is no new version under way, the main folder will be empty.

Old versions of posted files

Folders for posted files include a subfolder for old versions. It contains posted files that have been superseded by newer versions. For instance, if the main folder contains version 3, versions 1 and 2 (both the PDFs and the Word files) will be in the 'Old versions' subfolder.

Supporting files

In the MIGNEX collaboration folder, folders for deliverable drafts (e.g. D022-drafts) can include a folder for supporting files ('D022-supporting-files') containing related files that are not drafts of the deliverable itself, but, for instance, excel sheets for graphs or documents with text for specific sections. The purpose of this sub-folder is to avoid cluttering the main folder for the deliverable with files other than the current draft of the deliverable itself.

Task folders and deliverable folders

The relationship between tasks and deliverables varies. For instance, some tasks feeds into several deliverables, and others feed into one. In some cases, the task is essentially identical to writing the deliverable, while in other cases extensive analyses (e.g. in Stata) are needed. These differences should inform how files are divided between task folders (e.g. T084) and deliverable folders (e.g. D082).

Participant IDs

Survey respondents, focus group participants, key informants and policy experts are collectively referred to as *participants* in the project. Each participant has a unique ID, which allows for We must obtain informed consent from every participant. The general format is as follows:

[Country or research area ID]-[Data collection type]-[Other grouping, optional]-[sequential number]

Table 10 presents the construction of participant IDs for each data type.

Table 10. Construction of participant IDs

Data collection mode	Format for participant IDs
Survey	*To be determined*
Focus group	[Research area ID][focus group letter][1-digit Informant number] CPV1A3
Key informant interview	[Research area ID][2-digit Informant number] CPV104, CPV214
Policy expert interview	[Country ID][P for Policy][2-digit Informant number] CPVP01, AFGP03, GHAP12

- *Follow-up by Jessica Hagen-Zanker: determine format of survey respondent IDs in light of CAPI software capabilities.*

References

Bierer, B.E., Crosas, M. and Pierce, H.H. (2017) 'Data Authorship as an Incentive to Data Sharing.' *New England Journal of Medicine*, 376(17):1684-1687.

Long, J.S. (2009) *The workflow of data analysis using Stata*. College Station: Stata Press.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship.' *Scientific Data*, 3:160018.